

SIEBERTSTRASSE 4 · 81675 MÜNCHEN

TEL.: +49-89-4130 40 · FAX: +49-89-4130 4111 · FAX (Marken - Trademarks): +49-89-4130 44 00

Max-Planck-Gesellschaft zur
Förderung der Wissenschaften e.V.
Our. ref.: B 3366 PCT

Method for the identification of nucleic acid by means of electrospray mass spectrometry

The present invention relates to a method for detecting a nucleotide sequence in a nucleic acid molecule by means of predetermined probes with different masses each using electrospray mass spectrometry. Advantageously, the method of the invention allows at the same time for the characterisation of a variety of unknown nucleic acid molecules with a set of different probes. Moreover, this invention relates to a kit containing the probes and/or a probe support, optionally with nucleic acid molecules attached thereto.

The exact characterisation of nucleic acids is very time-consuming and expensive. It is possible to characterise an unknown DNA by sequencing it. This is the most precise way to analyse DNA. Sequencing of DNA, however, is very time-consuming and only necessary if the whole sequence is of interest. Only very short DNA segments (<1000 nucleobases) may be sequenced in one working process. If DNA fragments of more than 1000 nucleobases in length are to be sequenced on a large scale, it is necessary to subdivide the DNA, which makes the method more expensive.

Many statements, however, can also be made with a lower resolution. Yet, the methods that have been described up to now have the disadvantage that it might be necessary to use radioactivity and that only one probe can be used in one analysis. Such a method according to the state of the art, for instance, involves searching for partial information with an array of different target DNAs. An array of many thousands of target DNAs may be immobilized on a solid phase and subsequently, all target DNAs may be analysed together for the presence of a sequence by means of a probe (nucleic acid with a complementary sequence)^{1,2}. A match of the target DNA with the probe can be proven by hybridizing the two nucleic acids. Probes may be any nucleic acid sequences of different lengths. There are different methods for selecting optimal libraries of probe sequences that overlap only to a very small extent^{3,4}. It is also possible to compile the probe sequences purposefully in order to find certain target DNA sequences. Oligofingerprinting is an approach which uses this

technology. A library of target DNAs is scanned with short nucleic acid probes, with the probes mostly having a length of only 8-12 bases. One probe each is hybridized at once to a target DNA library that is immobilized on a nylon membrane. The probe is radioactively labelled and the hybridization is estimated according to the localisation of the radioactivity. For scanning an immobilized DNA array, fluorescently marked probes have also been used⁵. A similar method has been used for multiplexing the sequencing of DNA^{6,7}. Use is made of various vector systems for cloning target sequences. One clone each is pooled by each cloning vector, the sequencing reaction is carried out and the fragments are separated on the gel and the gel is blotted onto a nylon membrane. The different sequences of the cloning system are subsequently hybridized to the immobilized DNA so that the sequence belonging to the respective cloning system is obtained. Hereby, the scanning of the cloning system can also be carried out by means of a probe which is detectable by mass spectrometry⁸.

Any molecules capable of interacting sequence-specifically with a target DNA can be used as probes. Oligodeoxyribonucleotides are most commonly used. For this purpose, however, any modification of nucleic acids, e.g. peptide nucleic acids (PNA)^{9,10}, phosphorothioate oligonucleotides or methyl phosphonate oligonucleotides, are suitable. The specificity of a probe is extremely important. Phosphothioate oligonucleotides are not particularly preferred as their structure is modified by the sulphur atoms, which has a negative influence on the nature of the hybridization. This can be due to the fact that phosphothioate oligonucleotides are usually not synthesized free of diasteromers. In the past, there has been a similar purity problem with methyl phosphonate oligonucleotides, but these oligonucleotides are increasingly synthesized free of diasteromers. An essential difference between methyl phosphonate oligonucleotides and phosphorothioate oligonucleotides is the uncharged backbone of the former, which leads to a reduced hybridization dependency on buffer salts and, all in all, to a higher affinity due to reduced rejection. Peptide nucleic acids also have an uncharged backbone which, at the same time, drastically deviates in its chemical properties from the common sugar phosphate structure of the backbone of nucleic acids. The backbone of a PNA has an amide sequence instead of a sugar phosphate backbone of normal DNA. PNA hybridizes very well to a sequence complementary to DNA. The melting temperature of a PNA/DNA hybrid is higher than the one of the corresponding DNA/DNA hybrid and the hybridization dependency on buffer salts is, again, relatively small. Electrospray mass spectrometry (ESI) is a new, very efficient method for analysing biomolecules (Meng, C.K., Mann, M. und Fenn, J.B., Z. Phys. D. 1988, 10, 361-368; Siuzdak, G. Proc. Natl. Acad. Sci. USA, 1994, 91, 11290-11297). The basic features of this

method are that a solution of a mixture of biomolecules is conducted through a metallic needle to which a high voltage is applied (typically several kilovolts), into the vacuum of a mass spectrometer. Due to the electric instability and the decreasing pressure, the solution disperses in small drips, each carrying electric charge on their surfaces. In several skimming steps, the drips are transferred to the high vacuum of the mass spectrometer. Thereby, the solvent of the drips is gradually lost and the naked, multiply charged biomolecules are located in an unfragmented form in the high vacuum of the mass spectrometer. For these methods, there are several possibilities as regards the step which comprises the mass spectrometrical analysis. One possibility relates to the separation of molecules with different mass and charge by means of a magnate (sector, quadripole); another one relates to the analysis by means of measuring the time of flight. Hereby, the beam potential is pulsed vertically to the ion beam and the size of the particles is determined according to the time of flight. With these two embodiments it is possible to achieve an isotope resolution in the mass range of several thousand daltons. With new nanospray sources excellent spectrums can be recorded using only a few 100 fmol material.

Up to now, the ESI has essentially been successful for the analysis of peptides (Siuzdak, G., Proc. Natl. Acad. Sci. USA, 1994, 91, 11290-11297) and in particular for the analysis of peptide libraries produced combinatorially it has proven to be successful (Metzger, J.W., Kepter, C., Wiesmuller, K.H. and Jung, G., Anal. Biochem., 1994, 219, 261-277; Winger, B.E. and Campana, J.E., Rapid Commun. Mass Spectrom., 1996, 10, 1811-1813). The analysis of nucleic acids and modified variants has so far not been taken into consideration very often (Reddy, D.M, Rieger, R.A., Torres, M.C., and Iden, C.R., Anal. Biochem. 1994, 220, 200-207; Potier, N., Van Dorsselaer, A., Cordier, Y. Roch, O. and Bischoff, R., Nucleic Acids Res., 1994, 22, 3895-3903; Greig, M. and Griffey, R.H., Rapid Commun. Mass Spectrom., 1995, 9, 97-102). Thereby, problems arose as for example the fact that nucleic acid molecules are polyanions, which triggers additional problems in the ionisation.

Combinatorial syntheses¹³, i.e. the production of substance libraries starting from a mixture of preliminary stages are conducted in a liquid as well as in a solid phase. It is mainly the combinatorial solid phase synthesis that has established itself at a very early stage since the separation of by-products is particularly easy in this case. Only the target bonds attached to supports are retained in a washing step and at the end of the synthesis, they are isolated by specifically cleaving a linker. This technique allows in a very simple way for the simultaneous synthesis of a variety of different linkings to a solid phase and thus for obtaining chemically "pure" substance libraries^{14,15,16}. Consequently, chemistry can easily access the classes of linkings which are also synthesised to a solid phase in combinatorial, conventional syntheses

and thus, these are widely used. This mainly applies to peptide, nucleic acid and PNA libraries.

The synthesis of peptides is carried out by attaching the first N-protected amino acid (e.g. Boc) to a support, by a subsequent removal of the protection and reaction of the second amino acid to the free NH₂ group of the first one. In a further "capping" step, non-reacted amino functions are subjected to a further reaction in the next synthesis cycle. The protecting group at the amino function of the second amino acid is removed and the next building block may be coupled to it. For the synthesis of peptide libraries a mixture of amino acids is used in one or more steps. The synthesis of PNA and PNA libraries is carried out accordingly.

Nucleic acid libraries are mostly obtained by means of solid phase synthesis with mixtures of various phosphoamidite nucleosides. This may be conducted on commercially available DNA synthesizers without modifications in the synthesis protocols. Furthermore, various studies on the combinatorial synthesis of PNA libraries have been published^{17,18}. These studies deal with the structure of combinatorial sequences, i.e. the synthesis of PNAs in which single, specific bases of a sequence are replaced by degenerated bases and thus, an incidental sequence variation is obtained. The use of mass spectrometrical methods for analysing combinatorial libraries has also been described several times^{19,20,21,22}.

There are various methods for immobilizing DNA. The best-known technique is the solid linking of a DNA, which has been functionalized with biotin, to a streptavidin-coated surface²³. The affinity of this system corresponds to a covalent chemical bond without being one. In order to be able to covalently bind a target DNA to a chemically prepared surface, a certain functionality of the target DNA is required. DNA itself does not have any functionalization that would be suitable for this purpose. There are several possibilities for introducing an appropriate functionalization into a target DNA: two functionalizations that are easy to work with are primary, aliphatic amines and thiols. Such amines quantitatively react with N-hydroxy succinimide esters. Under suitable conditions, thiols quantitatively react with alkyl iodides. One difficulty is the introduction of such a functionalization into a DNA. The simplest way is to introduce it by means of a primer of a PCR. In said methods a 5'-modified primer (NH₂ and SH) and a bifunctional linker are used^{24,25,26}.

As regards the immobilization on a surface, it is mainly its nature that is of essential importance. Systems that have been described so far mainly consist of silicon or metal (magnetic beads). A further method for binding a target DNA is based on the

use of a short recognition sequence (e.g. 20 bases) in the target DNA for the hybridization to a surface-immobilized oligonucleotide²⁷. Enzymatic techniques for introducing chemically active positions into target DNA have also been described²⁸. In this case, a 5'-NH₂ functionalization is carried out enzymatically at a target DNA.

As has been described before, there is a number of methods known in the state of the art, which are all geared to an exact analysis of nucleic acids. Normally, these techniques are either very time-consuming and/or expensive.

Thus, the technical problem underlying the present invention was to provide a fast and cost-efficient method for the identification of target nucleic acids.

This technical problem has been solved by providing the embodiments characterised in the claims.

Therefore, the present invention relates to a method for detecting a nucleotide sequence in a nucleic acid molecule comprising the following steps:

- (a) hybridization of nucleic acid molecules to a set of probes of different nucleobase sequences, wherein each probe has a mass that differs from the one of all the other probes;
- (b) separation of the probes that were not hybridized;
- (c) detachment of a specifically hybridized probe in a solvent;
- (d) analysis of the hybridized probes in a solution by means of electrospray mass spectrometry; and
- (e) determination of the nucleic acid molecules by means of the probes hybridized to them.

Preferably, the nucleic acid molecules are fixed to a support. In this case, it is preferred that the positions of the probes on the probe support allow for an allocation to the nucleic acid molecule hybridizing therewith. The term "probe support", as used according to the invention, means any form of support material.

The method according to the invention (cf. Figure 1) advantageously combines methods for analysing target nucleic acids (oligofingerprinting), and the mass spectrometrical analysis of nucleic acids and modified nucleic acids. Thereby, a large number of different probes is used which allow for the detection of one or more nucleotide sequences in one nucleic acid molecule. So far, it has not been possible to combine these two methods, because the differences in the mass of the probes

did not allow for clear conclusions as to the sequence, and because the sensitivity of a mass spectrometric analysis of nucleic acids was not adapted to the probe amounts of an oligofingerprinting experiment. In the broadest sense of the invention, a nucleic acid molecule containing the nucleotide sequence(s) that is (are) (essentially) complementary is determined by the hybridization of a specific probe (or several specific probes). Due to this determination, it can also be proven that a specific nucleotide sequence or very similar sequences (for example in the case of a hybridization under non-stringent conditions, see below) is/are contained in the molecule. It is clear to the person skilled in the art that it is not always possible to exactly determine a nucleotide sequence searched for by means of a hybridization method since – even under stringent hybridization conditions – a hybridization of a probe might possibly take place in spite of so-called "mismatches" (for example from a certain minimum length of a probe or with positioning of the mismatch(es) that can be tolerated during hybridization). With a given nucleotide sequence of a probe, a complementary sequence in the nucleic acid molecule can only be determined with a certain estimation in some parts of the embodiments since, apart from exact complementary sequences, possibly also such sequences can be determined which are not exactly complementary in their sequences. Thus, the nucleotide sequence also comprises homologous nucleotide sequences which preferably exhibit a degree of homology of more than 90 %, particularly preferred more than 95 %. The present invention comprises all the above-mentioned embodiments.

Depending on which hybridization conditions are selected, the method of the invention may be used for detecting either specific nucleotide sequences or groups of nucleotide sequences having similar sequences. If, for instance, stringent hybridization conditions are selected, the probes used are only able to hybridize to the nucleotide sequences that are exactly complementary to their nucleobase sequences. If, however, non-stringent hybridization conditions are chosen, the probes used can detect all the nucleotide sequences that differ from the nucleobase sequences in that they still allow for hybridization under the conditions selected. In this way, the method according to the invention may also be used for detecting homologues, variants or alleles or a specific sequence. The person skilled in the art knows what stringent and non-stringent conditions are; cf. Sambrook et al., "A Laboratory Manual", CSH Press, Cold Spring Harbor, 2nd ed. 1989, Hames and Higgins (editors) "Nucleic Acid Hybridization, A Practical Approach", IRL Press, Oxford 1985. Stringent hybridization conditions are, for example, hybridization in 6 x SSC, 5 x Denhardt's reagent, 0.5 % SDS and 100 µg/ml denatured DNA at 65°C and washing in 0.1 x SSC, 0.1 % SDS at 65°C. Non-stringent hybridization conditions differ from the above-mentioned conditions insofar as, for example, hybridization and

washing are carried out at a lower temperature, e.g. at 50°C or 55°C and/or the amount of SSC is raised, for example, to 1 x or 2 x SSC.

The method of the invention also allows for the detection of several different sequences in one target DNA, wherein the different sequences may be complementary to different probes. At the very best, e.g. when using probes with overlapping sequences, the whole nucleotide sequence of a target sequence may be detected or clarified.

With the method of the invention it may first be determined whether a target nucleic acid was fixed to an immobilization support, which exhibits a sequence that can be hybridized to a probe under the conditions selected. If this is the case, the nucleic acid of the sample may be further analysed and characterised. As in the method of the invention, usually only a very small percentage of a sample has to be used for the analysis according to the invention, the nucleic acid that is not used may be further analysed with standard methods, e.g. sequencing methods.

The removal of the specifically hybridized probes in a solvent is preferably conducted by means of thermal denaturation or alkaline denaturation. In particular, aqueous solutions or mixtures of water and organic solvents are suitable solvents, provided the organic solvent can be mixed with water (e.g. water and acetonitrile). Preferably, the mixtures are used at a ratio of 1:1. As the hybridized molecules are to be denatured, the selection of the mixture, in addition, depends on the kind of molecules that are to be hybridized.

The method according to the invention may also be carried out several times either simultaneously or consecutively, with hybridization conditions being varied. In this way, it is possible, for example, to find out at first, whether a target DNA exhibits a specific degree of homology to other target DNAs, before searching for specific sequences.

The allocation of the probe hybridized at a specific position on the probe support to the immobilized sample is preferably conducted by means of a data processing system which allocates the respective spectrum recorded at a position of the probe support to a target DNA located exactly at the same position. Preferably, the target nucleic acids are located in a specific order on the surface or on the protein support.

In a preferred embodiment of the method according to the invention, the nucleic acid molecules are transferred to the surface of a support before or after step (a). For this

purpose, use can be made of a "magnetic bead", for example, which is functionalized in such a way that a target DNA functionalized correspondingly can be attached to it, or of a functionalization, for example with a microtitre plate in the different wells of which target DNAs may be immobilized. In this way, hybridization steps involving probe libraries and subsequent washing steps may be simplified. Biotin-streptavidin, NH₂-N-hydroxysuccinimidyl ester or SH-alkylhalide may be used.

In another preferred embodiment of the method of the invention, immobilization of the nucleic acid molecules is carried out at the surface of a probe support through a NH₂-, an epoxy- or a SH-function, through a protein-substrate, protein-protein or a protein-nucleic acid interaction by means of coating the surface of the probe support with a silicate or silane, or through an interaction of two hydrophobic building blocks.

If the probe support is coated with gold, coupling of the target DNA may take place by means of SH- or NH₂-functions introduced during the molecular biological preparation of the target DNA. There also is the reverse possibility of attaching a correspondingly modified DNA to functionalized gold particles. The company Nanoprobes Inc., Stony Brook, NY, for instance, sells gold-nanoparticles coupled with streptavidin or amino functions. By means of an amino functionalization, a subsequent coupling of the target DNA with a glass surface may be achieved. The target DNA is attached to a bifunctional linker (e.g. SIAB, Pierce Chemical, Rockford, IL, USA) through an SH-functionalization. Another possibility is to coat the surface directly with trimethoxy-3-amino propylsilane. Then, a target DNA may be coupled with the amino function through a bifunctional linker as described before.

In a particularly preferred embodiment, the protein-substrate interaction is a biotin-streptavidin or an antibody-antigen bond. Many probe supports usually consist of material to which neither proteins nor DNA molecules may be immobilized without further modification. One possibility of immobilizing is to coat the surface with gold since SH-functions, for example, may be coupled with it. For coupling, use can be made of bifunctional linkers having an SH-function and another function which is adapted to the functionalization of the target DNA. If the target DNA is, for example, biotin-functionalized, it should be possible to couple the linker with streptavidin. If the target DNA is NH₂-functionalized, the linker may have an N-hydroxysuccinimidyl ester-function.

In another particularly preferred embodiment, the protein-nucleic acid interaction is an attachment of the nucleic acid to Gene32, a protein which binds single-stranded DNAs in a way non-specific to sequences.

In another preferred embodiment of the method of the invention, the probes used are nucleic acids having a mass tag. According to this embodiment, the probes may also have several tags which may be located at different positions, e.g. at the 5' and the 3' end. The versatility and sensitivity of the method of the invention may be increased considerably by combining the amount and the localisation of mass tags, optionally in combination with charge tags.

In a particularly preferred embodiment the mass tags are at the same time charge tags, while in another particularly preferred embodiment, the nucleic acids additionally have a charge tag (cf. Figure 2).

Charge tagging may be conducted according to the method of Gut et al.^{11,12}. An amino-functionalized substrate (1 mM) is put in trimethylamine/CO₂ buffer (pH = 8.5, 200 mM) on ice at 0°C with 1 % ω -trimethylammoniumhexaneacid-N-hydroxysuccinimidylester (CT). After 30 minutes, the volatile buffer and the solvent in the vacuum are removed. The amino-functionalized substrate may, for example, be a combinatorially produced library of probes differing in their mass. By varying length and functionalization of the CT, the masses of the substrate library can be altered by a defined value (Fig. 2). As in the combinatorial synthesis, a probe library containing 64 probes differing in their mass in a range of 200 Da may be generated, the mass/charge tags increase the mass by units of 200 Da each, i.e. the first combinatorial synthesis preparation is generated with the smallest charge tag possible, the second with a mass/charge tag that has 200 Da more in weight, the third mass/charge tag that has another 200 Da more in weight, etc. In theory, this range can be extended at will as long as the mass spectrometer used is able to eliminate the difference between the two probes with neighbouring masses and as long as the synthesis seems to be feasible in practice. For probes with 10 nucleobases, a basic mass in the range of 2600 to 2800 Da is obtained. For mass spectrometers which can be bought at present, the range of mass that can be used and that has a sufficient accuracy as regards mass is below 4000 Da. Thus, seven ensembles of 64 probes can be used (all in all 448 probes).

The synthesis of peptides on an automated synthesizer takes place from the C-terminal end to the N-terminal end, the synthesis of nucleic acids from the 3' to the 5' end. There is the possibility of linking a primary amino function to one or to both ends in order to obtain mass shifting by means of one or two functionalizations. Alternatively, the masses of a library of combinatorially produced probes may also be altered by inserting some building blocks with a defined mass (e.g. amino acids in a combinatorial synthesis of PNAs) before inserting the combinatorial building blocks.

The first combinatorial synthesis starts directly at the support. For the second combinatorial synthesis, first of all two valines, for example, are coupled. Valine has a mass of 99 Da. With two valines the mass of the second combinatorial synthesis is changed by 198 Da compared to the first one. For the third combinatorial synthesis, first of all four valines are coupled, as a consequence, the mass of this ensemble is by 396 Da higher than the first one, etc. If a charge tag is necessary, it may be attached later on to the N-terminal end according to the above-mentioned method. It is also possible to couple the charge tags to the solid phase first and then to continue with the combinatorial synthesis (cf. for example Fig. 3).

Another possibility is to attach several fixed charges of the same polarity at the same time. Thus, the range of detection can be reduced. In a mass spectrometer, a molecule with two charges with the corresponding half of the mass is observed. This may be of advantage when the resolution of the mass spectrometer decreases drastically towards larger mass. It has been described before how to solidly link positive fixed charges to a probe. Negative fixed charges may have similar advantages. Positive and negative fixed charges can be generated with similar methods.

For the localisation, the amount and the combinatorics of the charge tags, the statements made in connection with the mass tags apply accordingly. Therefore, the sensitivity of an analysis may be improved by means of charge tagging at the 5' or 3' end of the library. This is particularly true if otherwise a neutral charge is generated by introducing alkylphosphonate groups into the non-randomized positions (cf. below). In addition, by means of mass tagging, which is inevitably related to charge tagging, the analysis of the fragments resulting from the cleavage at the phosphorothioate groups is facilitated. It is also possible to use phosphoroselenoates for achieving a preferred cleavage at certain positions.

In another preferred embodiment of the method of the invention, the probes are modified nucleic acid molecules.

In a particularly preferred embodiment, these modified nucleic acid molecules are PNAs, alkylated phosphorothioate nucleic acids or alkylphosphonate nucleic acids.

In another preferred embodiment, the probes of the method of the invention are generated by means of combinatorial solid phase synthesis. This, for example, can be carried out by means of the commonly used Boc solid phase synthesis which is also used commercially for the synthesis of PNA pure substances. In this case, all

four or more than one of the bases adenine, guanine, cytosine or thymine are incorporated next to each other in selected positions within a sequence of the probe DNA that is otherwise determined. This takes place by using not only one but several synthesis building blocks in one step of the solid phase synthesis. By varying several of these positions, a PNA library is created. After the PNA synthesis a charge tagging may be conducted at the free terminal amino function. This improves the dynamic range of the PNA analysis by means of ESI.

Therefore, in a preferred embodiment, the different building blocks for the solid phase synthesis are labelled in such a way that their mass or the mass of the probes synthesized from them can be differentiated in the mass spectrometer. This labelling is accomplished by introducing a different mass modification at the backbone in a way corresponding to the base. In this way, the synthesized probes obtain a mass that is specific to their sequences. If the probes now specifically bind to an immobilized target DNA, the information on the mass accessible in the mass spectrometrical experiment makes it possible to draw conclusions as to the sequences of the respective hybridized probe.

In another particularly preferred embodiment of the method of the invention, the building blocks are labelled with a methyl, ethyl, propyl, a branched or non-branched alkyl, a halogen-substituted branched or non-branched alkyl- or alkoxyalkyl, alkylaryl, arylalkyl, alkoxy or aryloxyalkyl group or with one of their deuterated or other isotopic variants. For, if building blocks are used the mass of which has not been modified, the molecular mass only makes it possible to draw conclusions as to the composition of the bases, but not as to the sequence, mass-labelled building blocks (i.e. PNA monomers substituted at the backbone) are used which are characteristic to any randomized position within the PNA library. As in this way the synthesis building block of a specific base has a different mass in position X than in position y, it is possible to differentiate also different sequences with the same gross composition of the bases through the molecular mass. The respective substituents are selected by numeric values in such a way that a determined mass corresponds unequivocally to any possible sequence due to the described characteristics of the library synthesis.

The synthesis of nucleic acid probe libraries is conducted at the synthesizer by inserting mixtures of different nucleoside derivatives (usually phosphoramidites) at the positions to be randomized. The libraries resulting therefrom can also be used as probes in the above-described method. In order to differentiate the different sequences within the library, a specific bond breaking at the phosphodiester bonds located at a determined side of the randomized positions is to be carried out.

Therefore, in another preferred embodiment of the method of the invention, the probes have at least one modification in a defined position away from the randomized nucleotides which allow for the cleavage of the probe.

In a particularly preferred embodiment, this modification is the introduction of a phosphorothioate group and/or an RNA base and/or a phosphodiester bond.

If a probe contains three randomized positions, at least two such bond breakings are necessary (Fig. 6). As a result, three fragments are obtained which contain one randomized position each and which thus allow for a direct conclusion as to the variable base through its mass due to the composition otherwise known. The cleavage of the described bonds may also be conducted incompletely. This makes it possible to include larger fragments to secure the sequence information or to concretize it in the case of ambiguities. The specific bond breaking at the randomized positions is carried out by introducing phosphorothioate groups already during the synthesis of the library. These may first be hydroxyalkylated with hydroxyalkylhalides and then cleaved selectively under alkaline conditions. Alternatively, the samples may be set up in such a way that uracil is inserted next to a randomized position. The backbone can then be broken at the same position by means of uracil DNA glycosylase and subsequent alkaline treatment.

The above-described principle of mass tagging for certain components in determined positions may be used for various partial libraries which then may be united to a larger library.

Therefore, in another preferred embodiment, the probes are generated as partial libraries having different masses and/or charge tags. It is necessary to mass-label also the partial libraries in their syntheses so that specific analysed masses allow for conclusions as to that specific partial library. Usefully, this is carried out through natural amino acids which are easy to attach to a PNA library in a solid phase synthesis. By means of mass-labelling different partial libraries, it is possible to considerably extend the range of mass which is disposable for the analysis of the total library.

Moreover, the present invention relates to a kit containing the probes and/or parts of an immobilization system (probe supports), optionally with the nucleic acid molecules bound. (Coated) "magnetic beads" or a (coated) microtitre plate are suitable immobilization systems. As has been described before, the surface of the

immobilization system is pretreated and thus allows for the binding of the nucleic acids. Preferably, this pretreatment is carried out in a chemical way.

In this application, a number of publications are cited which are herewith incorporated into the specification by reference.

The Figures show:

Figure 1

Scheme of fingerprinting with mass spectrometrical scanning.

1. Combinatorially generated library of probes that can be differentiated by their mass.
2. The library is hybridized to an immobilized target DNA. After washing it intensively (to remove non-specifically bound probes), the correctly hybridized probes are separated from the target DNA.
3. This solution is analysed in an electrospray mass spectrometer. The partial sequences of the target DNA are determined by clearly allocating the mass to the sequences.

Figure 2

N-terminal mass/charge tagging.

A charge tag may be introduced at the 5' of a nucleic acid or a modified version of a nucleic acid, or at the N-terminal end of a PNA. The charge tag has an N-hydroxy-succinimidester function and a quarternary ammonium group. These two functionalizations are separated by one group R₁. R₁ is used to vary the mass of the charge tags (coupling of the charge tags is carried out at a slightly alkaline pH (8.5) in an aqueous solution on ice. The reaction is completed within 30 minutes).

Figure 3

Principle of synthesis of PNA libraries using syntheses of mass-labelled syntheses that are carried out in parallel as well as mass-labelled building blocks L (randomized position). Various bases with different substituents each are used (as mass tags) in

the randomized positions at the backbone. The mass of the respective PNA molecule is allocated unequivocally to its sequence.

Figure 4

Calculated mass spectrum of a PNA library.

Two different solid phase syntheses (one of which is mass-labelled) with 32 different sequences each make up 64 different mass peaks, each of them being allocated to a specific sequence from a PNA library with 3 variable positions with four bases (A, C, G and T being inserted). The calculation is based on the substituents listed in Table 1. For the calculations the computer programme MASP (© Dr. Christoph Steinbeck) was used.

Figure 5

Possible synthesis components for the combinatorial Boc solid phase synthesis as are also used for the synthesis of PNA libraries that can be differentiated due to their mass. The components listed in Table 1 are shown.

Figure 6

Sequence analysis of probes by means of specific bond breakings.

The different sequences of a library of DNA probes can also be identified through mass spectrometry by means of specific bond breakings at the randomized positions (N = A, G, C or T). A phosphorothioate function at which the DNA may be cleaved specifically (e.g. with idoethanol) is inserted during the solid phase synthesis. This is shown schematically without taking the exact chemical nature of the fragments into account. Since the sequence is known to a great extent, the mass of the fragments makes it possible to determine the total sequence without having to carry out a complete sequencing. The substituent R_1 stands for any possible further DNA sequence, R_2 for another possible sequence or $-H$. R_3 stands for $-OH$ (phosphodiester) or alkyl- (alkylphosphonate).

Example 1: Explanation of an embodiment of the whole method of the invention

A target DNA is generated with a function allowing for its immobilization at a surface. The target DNA is produced by means of PCR and in said PCR, a primer functionalized with biotin is used. For the immobilization, the surface is coated with streptavidin. A library of probes that can be differentiated due to their mass is hybridized to a target DNA in the above-described way. Complementary building blocks of the probe library bind to the respective target DNAs. Then, they are washed thoroughly to remove non-specifically bound probes. The specific probes are separated from the target DNA by means of heat denaturation. The solution is analysed in an electrospray mass spectrometer. Due to the mass found and the fact that the mass of the probes is distinguishable conclusions as to partial sequences in the target DNA can be drawn.

Example 2: Analysis of target DNA immobilized via a biotin-streptavidin bridge

A target DNA is attached to a streptavidin-coated magnetic bead by means of a biotin function. A library of probes distinguishable by mass is hybridized to target DNA. Non-specific hybridizations are washed off. By heating in an aqueous solution, the specific hybridized probes are detached. The aqueous solution is analysed in an electrospray mass spectrometer. Due to the mass found and the fact that the mass of the probes is distinguishable conclusions as to partial sequences in the target DNA can be drawn.

Example 3: Immobilization of target DNA

In a preferred variant of the method, the target DNA is attached to a bead consisting of an epoxy-functionalized acrylpolymer. Subsequently, the epoxy functions that have not reacted are deactivated with a surplus of an amine. The example is continued as described in Example 2.

Example 4: Analysis of various target DNAs in a microtitre plate

In a microtitre plate 96 or 384 different biotin-functionalized target DNAs are immobilized in one well each. In each well, hybridization to a library of probes distinguishable by mass is carried out. The plate is washed in order to remove non-specifically bound hybridizations. The specific hybridizations are solved in the respective well by heating in an aqueous solution. The aqueous solution in each well is analysed in an electrospray mass spectrometer. Due to the mass found and the fact that the mass of the probes is distinguishable allow for conclusions as to partial sequences of the target DNA in the respective well.

Example 5: Synthesis of a PNA library with 64 different sequences

Figure 4 and Table 1 show an example of a PNA library with a clear relation between mass and sequence. In an otherwise random PNA, three positions are to be varied in the combinatorial solid phase synthesis. With 4 different bases this corresponds to 64 possible linkings. In the present example, two separate syntheses are conducted with the building blocks listed in Table 1. The second synthesis is additionally mass-labelled by adding two valine units. By each synthesis 32 compounds are obtained which can all be differentiated by their mass. Both partial libraries can be united to one library containing 64 different sequences, each one with a specific mass.

The 64 peaks do not overlap, each peak corresponds to a specific sequence from a PNA library with three randomized base positions (4 bases each).

Example 6: Creation of a probe library with 448 probes

7 syntheses are carried out according to the method described in the latter example and subsequently, the mass/charge tags are coupled.

	mass range
1. 64 probes: charge tag- TCP ₁ GAP ₂ GAP ₃ G	2600-2800 Da
2. 64 probes: charge tag+200 Da mass tag- TCP ₁ AGP ₂ GAP ₃ G	2800-3000 Da
3. 64 probes: charge tag+400 Da mass tag- TCP ₁ AGP ₂ AGP ₃ G	3000-3200 Da
4. 64 probes: charge tag+600 Da mass tag- TCP ₁ AAP ₂ AGP ₃ G	3200-3400 Da
5. 64 probes: charge tag+800 Da mass tag- TCP ₁ AAP ₂ GAP ₃ G	3400-3600 Da
6. 64 probes: charge tag+1000 Da mass tag- TCP ₁ GAP ₂ GAP ₃ G	3600-3800 Da

7. 64 probes: charge tag+1200 Da mass tag- TCP₁GAP₂AGP₃G 3800-4000 Da

In the above-mentioned synthesis series, the sixth synthesis serves as an internal control. Alternatively, the following synthesis may also be conducted:

6. 64 probes: charge tag+1000 Da mass tag- TCP₁GGP₂GAP₃G 3600-3800 Da.

Example 7: Coating of a microtitre plate with protein

The surface of a microtitre plate is coated with Gene32, a protein that non-specifically binds single-stranded DNA. After coating the target with this protein, an array of target DNAs can be put onto it. If the array of the target DNAs consists of cDNA, these were primed with oligo-dT (e.g. dTTTTTTTTTTTTTT) in this PCR. Oligo-dT interacts heavily with Gene32. The covalent coupling of the oligo-dT to Gene32 may be achieved by means of photocrosslinking with short UV light^{29,30}. After this immobilization, a library of probes may be used for analysing the target DNA in the above-described way. Use may also be made of sequence-specific protein/DNA interaction (e.g. GCN4/AP1).

Example 8: Immobilization of target DNA at an epoxy-functionalized solid phase.

10 mg support (Eupergit C250 L, Röhm Pharma Polymere) are suspended in 1 ml immobilization buffer (1 M K₂HPO₄/KH₂PO₄, pH 7.5). 1.5 nmol of the target DNA that is to be immobilized is added and incubated for 24 hours at room temperature while shaking gently. The supernatant is removed and the epoxy functions which have not reacted are deactivated through treatment with 1 M glycine solution for 24 hours at room temperature. The supernatant is removed and the support is washed several times. In this way, the immobilized DNA can be stored for a longer time at -20°C.

Example 9: Selective hybridization with "charge tagged" PNAs

The immobilized target DNA is incubated for 15 min at 65°C with the library of "charge tagged" PNA probes (80 pmol in 20 µl/8 pmol immobilized DNA) in hybridization buffer (10 mM Tris/HCl, 5 mM NH₄Cl, 15 % formamide) and subsequently, the probes that have not been hybridized are removed by washing

three times with hybridization buffer. Then, the dehybridization is carried out in 40% acetonitril at 80°C. An aliquot of this solution is lyophilized, put into water and analysed in a ESI mass spectrometer.

base	position 1	position 2	position 3
A	H	<i>i</i> Pr	H
T	H	Me	<i>i</i> PrOCH ₂
C	H	H	<i>i</i> PrOCH ₂ *
G	H	<i>i</i> Bu	H*

Table 1: Substituents fixed at the PNA subunits for producing a mass-labelled library with an unequivocal mass/sequence.

Relation. *: second synthesis, mass-labelled with 2 valine units. Figure 5 shows the corresponding synthesis building blocks.

References

¹ Hoheisel, J.D. and Lehrach, H. 1993. Use of reference libraries and hybridsationfingerprint for relational genome analysis. FEBS. 325: 118-122.

² Scholler, P., Karger, A.E., Meier-Ewert, S., Lehrach, H., Delius, H. and Hoheisel, J.D. 1995. Fine-mapping of shotgun template-libraries; an efficient strategy for the systematic sequencing of genomic DNA. Nucleic Acids Res. 23: 3842-3849.

³ Brenner, S. 1994. Methods for sorting polynucleotides using oligonucleotide tags. US Patent. 5,604,097.

⁴ Brenner, S. 1995. Minimally cross-hybridizing sets of oligonucleotide tags. US Patent. 5,635,400.

⁵ Guo, Z., Guilfoyle, R.A., Thiel, A.J., Wang, R. and Smith, L.M. 1994. Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotides arrays of glass supports. Nucleic Acids Res. 22: 5456-5465.

⁶ Church, G.M. and Kieffer-Higgins, S. 1988. Multiplex DNA sequencing. Science. 240: 185-188.

⁷ Church, G.M. and Kieffer-Higgins, S. 1990. Multiplex analysis of DNA. US Patent. 5,149,625.

⁸ Arlinghaus, H.F., Kwoka, M.N., Guo, X.-Q. and Jacobson, K.B. 1997. Multiplex DNA sequencing and diagnostics by hybridization with enriched stable isotope labels. Anal. Chem. 69: 1510-1517.

⁹ Nielsen, P.E., Buchardt, O., Egholm, M. and Berg, R.H. 1993. Peptide nucleic acids. US Patent. 5,539,082.

- ¹⁰ Buchardt, O., Egholm, M., Berg, R.H. and Nielsen, P.E. 1993. Peptide nucleic acids and their potential applications in biotechnology. *Trends in Biotechnology*, 11: 384-386.
- ¹¹ Gut, I.G. and Beck, S. 1995. Method of nucleic acid analysis. Patent WO96/27681.
- ¹² Gut, I.G., Jeffery, W.A., Pappin, D.J.C. and Beck, S. 1997. Analysis of DNA by 'charge tagging' and matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* 11: 43-50.
- ¹³ Lowe, G. 1995. Combinatorial Chemistry. *Chem. Soc. Rev.* 24: 309.
- ¹⁴ Jung, G., Früchtel, J.S. 1996. Organische Chemie an fester Phase. *Angew. Chem.* 108: 19-46.
- ¹⁵ Choong, I.C., Ellman, J.A. 1996. Solid-Phase Synthesis: Applications to Combinatorial Libraries. *Ann. Reports in Med. Chem.* 31: 309.
- ¹⁶ Pirrung, M.C. 1997. Spatially Addressable Combinatorial Libraries. *Chem. Rev.* 97: 473.
- ¹⁷ Cook, P.D., Kiely, J. and Sprankle, K. 1994. Peptide nucleic acid combinatorial libraries and improved methods of synthesis. US Patent. 5,539,083.
- ¹⁸ Nielsen, P.E. Peptide nucleic acids: a new dimension to peptide libraries and aptamers. *Methods in Enzymology*. 426-433.
- ¹⁹ Metzger, J.W., Stevanovic, S., Brünjes, J., Wiesmüller, K.-H., Jung, G. 1994. Electrospray Mass Spectrometry and Multiple Sequence Analysis of Synthetic Peptide Libraries. *Methods: A Companion to Methods in Enzymology* 6: 425-431.
- ²⁰ Loo, J.A., DeJohn, D.E., Andrews, P.C. 1996. Application of Mass Spectrometry for Characterizing and Identifying Ligands from Combinatorial Libraries. *Ann. Reports in Med. Chem.* 31:319.

- ²¹ Pomerantz, S.C., McCloskey, J.A., Eaton, B.C., 1997. Deconvolution of Combinatorial Oligonucleotide Libraries by Electrospray Ionization Tandem Mass Spectrometry. *J. Am. Chem. Soc.* 119: 3861.
- ²² Carr, S.A., Benkovic, S.J., Winograd, N. 1996. Evaluation of Mass Spectrometric Methods Applicable to the Direct Analysis of Non-Peptide Bead-Bound Combinatorial Libraries. *Anal. Chem.* 68: 237.
- ²³ Uhlen, M. et al. 1988, *Nucleic Acids Res.* 16, 3025-3038.
- ²⁴ Chrisey, L.A., Lee, G.U. and O'Ferrall, C.E. 1996. Covalent attachment of synthetic DNA to self-assembled monolayer films. *Nucleic Acids Res.* 24: 3031-3039.
- ²⁵ Timofeev, E.N., Kochetkova, S.V., Mirzabekov, A.D. and Florentiev, V.L. 1996. Regioselective immobilization of short oligonucleotides to acrylic copolymer gels. *Nucleic Acids Res.* 24: 3142-3148.
- ²⁶ O'Donnell, M.J., Tang, K., Köster, H., Smith, C.L. and Cantor, C.R. 1997. High-density, covalent attachment of DNA to silicone wafers for analysis by MALDI-TOF mass spectrometry. *Anal. Chem.* 69: 2438-2443.
- ²⁷ Cantor, C.R. 1995. Methods of preparing probe array by hybridization. US Patent. 5,631,134.
- ²⁸ Bruick, R.K., Koppitz, M., Joyce, G.F. and Orgel, L.E. 1997. A simple procedure for constructing 5'-amino-terminated oligodeoxynucleotides in aqueous solution. *Nucleic Acids Res.* 25: 1309-1310.
- ²⁹ Hockensmith, J.W., Kubasek, W.L., Vorachek, W.R. and von Hippel, P.H. 1993. Laser cross-linking of proteins to nucleic acids. *J. Bio. Chem.* 268: 15712-15720.
- ³⁰ von Hippel, P.H. 1994. Protein-DNA recognition: New perspectives and underlying themes. *Science.* 263: 769-770.